Final Performance Report
Grant FA9550-07-1-0366

04/01/2007-11/30/2009

Steven I. Marcus, Michael C. Fu, and Jiaqiao Hu
January 25, 2010

Abstract

The researchers made significant progress in all of the proposed research areas. The first major task in the proposal involved simulation-based and sampling methods for global optimization. In support of this task, we have discovered two new innovative approaches to simulation-based global optimization; the first involves connections between stochastic approximation and our model reference approach to global optimization, while the second connects particle filtering and simulation-based approaches to global optimization. We have also made significant progress in population-based global optimal search methods, applications of these algorithms to problems in statistics and clinical trials, and efficient allocation of simulations.

In support of the second task, we have made progress incorporating simulation-based and sampling methods into Markov Decision Processes (MDPs). We have made significant progress on new sampling methods for MDPs, simulation-based approaches to partially observable Markov decision processes (POMDPs), and applications of these algorithms.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
25-01-2010	Final Report	1 April 2007- 30 Nov 2009

4. TITLE AND SUBTITLE
Final Performance Report
Grant FA9550-07-1-0366
January 25, 2010

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-07-1-0366

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Steven I. Marcus, Michael C. Fu, Jiaqiao Hu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Office of Research Administration
and Advancement
Rm. 2100 Lee Building
University of Maryland
College Park, MD 20742

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/PKC (ATTN: Wendy Veon)
875 North Randolph Street, Suite 325, Room 3112, Arlington, VA 22203-1768

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The researchers made significant progress in all of the proposed research areas. The first major task in the proposal involved simulation-based and sampling methods for global optimization. In support of this task, we have discovered two new innovative approaches to simulation-based global optimization; the first involves connections between stochastic approximation and our model reference approach to global optimization, while the second connects particle filtering and simulation-based approaches to global optimization. We have also made significant progress in population-based global optimal search methods, applications of these algorithms to problems in statistics and clinical trials, and efficient allocation of simulations.

In support of the second task, we have made progress incorporating simulation-based and sampling methods into Markov Decision Processes (MDPs). We have made significant progress on new sampling methods for MDPs, simulation-based approaches to partially observable Markov decision processes (POMDPs), and applications of these algorithms.

15. SUBJECT TERMS
Simulation-based optimization, global optimization, estimation, Markov Decision Processes, sampling

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**c. THIS PAGE** Unclassified		23	Steven I. Marcus
					19b. TELEPHONE NUMBER *(include area code)* 301-405-7589

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

1 Introduction

In this research project, we proposed to investigate basic questions aimed at challenges in information superiority, logistics, and planning for the Air Force of the future. In particular, we proposed to investigate simulation-based methodologies for global optimization and planning that can be effective tools in an integrated approach to Global Awareness (Intelligence, Surveillance and Reconnaissance, or ISR), Command and Control, planning, and logistics.

Such systems are exceedingly complex, and we combined four approaches in the study of such problems:

- Developing and studying efficient simulation-based and sampling methodologies for global optimization problems;
- Studying the application of these global optimization methodologies to practical problems, such as those arising in planning for unmanned aerial vehicles and data mining;
- Developing and studying efficient simulation-based and sampling methodologies for problems of dynamic decision making under uncertainty;
- Studying the application of these dynamic optimization methodologies to practical problems, such as inventory control, preventive maintenance, and optimal stopping.

1.1 Simulation-Based and Sampling Methods for Global Optimization

Simulation is used to model complex stochastic systems arising in applications from supply chain management to financial engineering to telecommunications, among many others. In addition to performance evaluation, optimization — or at least improvement — of the system is clearly a desirable objective.

Following [53], we distinguish between *instance-based* and *model-based* global optimization solution methods. In instance-based methods, the search for new candidate solutions depends directly on previously generated solutions, e.g., simulated annealing [33], genetic algorithms (GAs) [18], tabu search [17], and nested partitions [41]. On the other hand, in model-based algorithms, new candidate solutions are generated via an intermediate *probability model* that is iteratively updated. Our research has focused on the model-based optimization framework, which involves the following ingredients:

(0) specify probability distribution over solution space;
(I) generate candidate solutions by sampling from distribution;
(II) estimate performance of (and possibly improve) candidate solutions;
(III) update distribution based on selected ("elite") set of candidate solutions.

This approach retains the primary strengths of population-based approaches such as genetic algorithms — improving upon simulated annealing, which works with a single iterate at a time, while at the same time providing more flexibility in exploring the entire solution space, introducing more structure in the search procedure, and allowing theoretical properties to be studied regarding both finite-time performance and asymptotic convergence. The theory behind the framework is rigorous, but based on an idealized version of the last

three ingredients, specifically the distribution sequence, sampling from the distribution sequence (or from a surrogate or approximation), and estimation of the performance, since it is observed through simulation. Schematically, we seek a sequence of distributions

$$g_0, g_1, g_2, ... \longrightarrow g_\infty,$$

where g_∞ concentrates its mass around the optimal solutions.

Examples for the sequence of distributions $\{g_k\}$ include the following:

(a) proportional selection scheme — introduced in estimation of distribution algorithms (EDAs) [48], and the instantiation of our model reference adaptive search (MRAS) method in [25];

(b) Boltzmann distribution with decreasing (nonincreasing) temperature schedule — used in the annealing adaptive search (AAS) [45, 40];

(c) optimal importance sampling measure — used in the cross-entropy (CE) method [13].

Case (c) is easiest to implement, but may not converge to a global optimum. The other two cases have nice theoretical properties: Case (a) guarantees improvement in expectation [25], whereas case (b) guarantees improvement in stochastic order [45]. All three will be used in the proposed research.

However, in all three cases, the sequence $\{g_k\}$ is unknown explicitly a priori, or else the problem would essentially be solved. So at iteration k, sampling is done from a surrogate distribution(s) that approximates g_k. There are two main approaches that have been adopted:

- *Markov chain Monte Carlo* approximation to the target distribution at *each iteration* [45], which involves generating a *sequence* of sample points from a sequence of distributions $\{\pi_{ki}\}$ following a Markov chain that asymptotically converge to the target distribution (at that iteration), i.e.,

  $$\pi_{k1}, \pi_{k2}, ... \longrightarrow g_k;$$

  e.g., a common implementation is the "hit-and-run" algorithm [42, 46, 47];

- *projection* onto distributions that are easy to work with, e.g., use a family of *parameterized* distributions $\{f_\theta\}$, and project g_k onto the family to obtain a sequence that converges to the (final) target distribution, i.e.,

  $$f_{\theta_0}, f_{\theta_1}, f_{\theta_2}, ... \longrightarrow g_\infty;$$

  a common implementation minimizes the Kullback-Leibler (KL) divergence between f_{θ_k} and g_k at each iteration, because it leads to analytically tractable solutions if the parameterized distributions are from the exponential family.

The first approach is adopted by AAS, and generates a *sequence* of candidate solutions at each iteration. Our MRAS method and the CE method follow the second approach, which leads to a *population* of candidate solutions, from which an elite set is selected and used to update the distribution.

1.2 Simulation-Based and Sampling Methods for Markov Decision Processes

Simulation optimization problems arising in supply chain management, path planning for unmanned aerial vehicles, financial engineering, and telecommunications are characterized by two critical aspects: changing *dynamics* and *stochastic* events. For example, effective supply chain management requires optimal responsive actions in the face of both gradual shifts in demand patterns – e.g., due to technology advances – and sudden unpredictable disruptions in production capacity – e.g., due to an unanticipated manufacturing facility shutdown. Such systems often require computationally expensive *simulation* models for performance estimation, such as modeling the operations of an entire semiconductor fabrication facility, where simulation runtime is typically on the order of hours. Markov decision processes (MDPs) provide a powerful paradigm for modeling optimal decision making under uncertainty in these settings, but MDPs suffer from the well-known curse of dimensionality, which can include exponential growth in the size of state spaces and action spaces with the problem size; thus, direct numerical solution of MDPs for large-scale real-world problems *is not currently feasible*. In general, heuristics and approximations are employed to simplify the MDP model. Perhaps the most successful examples of this approach has been approximate dynamic programming using value function approximation (cf. [2, 12, 37, 39, 44]), but the size of problems that can be solved remains relatively small compared to large-scale real-world problems of interest. The approaches studied in this research are meant to complement these highly successful techniques.

2 Research Results

2.1 Simulation-Based and Randomized Methods for Global Optimization

Consider finding the optimal solution to the problem of the form:

$$x^* \in \arg\min_{x \in \mathbb{X}} H(x), \tag{1}$$

where $\mathbb{X} \subseteq \Re^n$ is the solution space, which can be either continuous or discrete, and $H(\cdot) : \mathbb{X} \rightarrow \Re$ is a bounded, deterministic function. We assume the existence of the optimal solution x^*. However, the function H is not necessarily convex or even continuous, and there could be multiple local minima. Note that in a more general stochastic setting, the objective function H itself may take the form of an expected value of a sample performance h, $H(x) = E[h(x, \psi)]$, where ψ is a random variable (possibly depending on x) representing the stochastic effects of the system, and only estimates of noisy function h are available.

The theoretical properties and practical efficiencies of model-based methods are primarily determined by the two key questions of how to update the probability models and how to efficiently generate candidate solutions from them. In [25], the PIs have introduced a general model-based randomized search framework called model reference adaptive search (MRAS), where these difficulties are circumvented by sampling candidate solutions from a family of *parameterized* distributions $\{f_\theta(\cdot), \theta \in \Theta\}$ (henceforth referred to as sampling distributions)

and using a sequence of intermediate *reference* distributions $\{g_k\}$ to facilitate and guide the updating of the parameters associated with the parameterized family. The idea of projecting to the exponential family of distributions was first introduced by the CE method; see [25] for the differences between MRAS and the CE method. The idea is that the parameterized family is specified with some structure so that once the parameter is determined, sampling from each of these distributions should be a relatively easy task. An additional advantage of using the parameterized family is that the task of updating the entire distribution simplifies to the task of updating its associated parameters. In MRAS, the parameter updating is carried out by minimizing the Kullback-Leibler (KL) divergence between the parameterized family and the reference distributions $\{g_k\}$.

The sequence $\{g_k\}$ is primarily used to express the desired properties of the method. Thus, these reference distributions are often selected so that they can be shown to converge to a degenerate distribution with all probability mass concentrated on the set of optimal solutions. Intuitively, the sampling distribution f_{θ_k} can be viewed as a compact approximation of the reference distribution g_k (i.e., the projection of the reference distribution onto the parameterized family). Thus, the hope is that the sequence of sampling distributions retain the convergence properties of the sequence of reference distributions $\{g_k\}$. The key steps in each iteration (k) of MRAS are the following:

1. Given parameter θ_k, generate a set of N candidate solutions $\Lambda_k = \{X_k^1, \ldots, X_k^N\}$ by sampling from $f_{\theta_k}(\cdot)$, and obtain their objective function values $H(x)\ \forall x \in \Lambda_k$.

2. Determine a set of "elite" candidate solutions $\Lambda_{elite} \subseteq \Lambda_k$ (e.g., the set of solutions with the best performance function values).

3. Update the parameter θ_{k+1} based on Λ_{elite} by minimizing the KL divergence:

$$\theta_{k+1} = \arg\min_{\theta \in \Theta} \mathscr{D}(g_{k+1}, f_\theta),$$

$$\mathscr{D}(g, f) := \int_{x \in \mathbb{X}} \ln \frac{g(x)}{f(x)} g(dx),$$

   and set $k \leftarrow k + 1$.

Step 2 above is in the same spirit as the selection scheme used in many population-based approaches such as genetic algorithms.

2.1.1 Model Reference Framework for Natural Exponential Families (NEFs)

For the exponential family of distributions, the minimization in step 3 of the basic algorithm can be carried out in analytical form, which makes MRAS also easy to implement efficiently.

Definition: *A parameterized family* $\{f_\theta(\cdot), \theta \in \Theta \subseteq \Re^m\}$ *on* $\mathbb{X}$ *is called a natural exponential family if there exist functions* $\Gamma : \Re^d \to \Re^m$ *and* $K : \Re^m \to \Re$ *such that* $f_\theta(x) = \exp\left(\theta^T \Gamma(x) - K(\theta)\right)$*, where* $K(\theta) = \ln \int_{\mathbb{X}} \exp(\theta^T \Gamma(x)) \nu(dx)$*, and* ν *is the Lebesgue/discrete measure on* $\mathbb{X}$*.*

The function $K(\theta)$ plays an important role in the theory of natural exponential families. It is strictly convex with $\nabla K(\theta) = E_\theta[\Gamma(X)]$ and Hessian matrix $\mathrm{Cov}_\theta[\Gamma(X)]$. Therefore, the

Jacobian of the *mean vector function* $m(\theta) := E_\theta[\Gamma(X)]$ is strictly positive definite and thus invertible. From the inverse function theorem, it follows that $m(\theta)$ is also invertible.

When NEFs are used in the framework with the sample size N adaptively increasing with rate $\gamma > 1$, the following result [25] establishes the convergence of the sequence of parameters $\{\theta_k\}$ to the optimal parameter θ^*.

Theorem 1 *Let $\beta > 0$ be a constant such that the set $\{x : S(H(x)) \geqslant \frac{1}{\beta}\}$ has a strictly positive Lebesgue/counting measure. If $\gamma > (\beta S(H(x^*)))^2$, then $\lim_{k\to\infty} \theta_k = m^{-1}(\Gamma(x^*))$ w.p.1.*

This theorem, when restricted to many special cases, implies the convergence of the sequence of sampling distributions $\{f_{\theta_k}(\cdot)\}$ to a degenerate distribution on the set of optimal solutions. For example, for continuous optimization problems, if the multivariate Gaussian densities $\mathcal{N}(\mu_k, \Sigma_k)$ with mean vector μ_k and covariance matrix Σ_k are used as parameterized distributions, then a simple interpretation of the theorem gives

$$\lim_{k\to\infty} \mu_k = x^* \quad \text{and} \quad \lim_{k\to\infty} \Sigma_k = 0 \quad \text{w.p.1.}$$

2.1.2 Extensions

The MRAS methodology is extended in [26, 24] to an algorithm called Stochastic Model Reference Adaptive Search (SMRAS) for finding the global optimal solution to stochastic optimization problems, for situations in which the objective function cannot be evaluated exactly, but can be estimated with some noise (e.g., via simulation). We prove that SMRAS converges asymptotically to a global optimal solution with probability one for both stochastic continuous and discrete problems. Numerical studies have been carried out to validate the method.

We have made the important discovery that there exists a close relationship between the model-reference framework and the well-known stochastic approximation method [28]. This relationship explains why these model-based algorithms work well for hard optimization problems with little structure – by implicitly converting and transforming the underlying problem into an equivalent problem on the parameter space with smooth differential structures – and any model-based algorithm that can be accommodated by the framework can essentially be viewed as a gradient-based recursion on the parameter space for solving the equivalent smoothed problem. This new discovery provides a unifying framework for analyzing the convergence and convergence rate behavior of model-based algorithms, in the sense that the available tools and results of stochastic approximation from over half a century can be applied to study model-based algorithms for general non-differentiable optimization problems. Moreover, this equivalence relationship will also help us to understand the capability and limitation of model-based algorithms, and provide insight into designing new algorithm instantiations. Our empirical implementation of model-based algorithms based on a pure gradient interpretation indicates significant performance improvement over their original versions. We believe that this new direction will eventually lead to robust and efficient new simulation and sampling-based techniques capable of handling complex optimization problem involving hundreds of decision variables.

Another important aspect we have worked on is the simulation/sampling efficiency issue in model-based algorithms. This is motivated by the fact that many model-based algorithms have a demanding computational requirement per iteration, since a sufficient number of candidate solutions need to be collected to update the probability models over the solution space. Efficient simulation and sampling procedures in model-based algorithms will help to capture reliable information in updating probability models with only a modest computational effort, and thus further enhance the value of such algorithms. Our approach is based on modeling the simulation and sampling process in a typical model-based algorithm as a sequential decision making problem, where the ultimate goal is to maximize algorithm performance subject to a given computational budget constraint. Some theoretical and empirical findings are contained in [22]. Our results show that for high-dimensional optimization problems, the proposed computing budget allocation scheme could yield orders of magnitude savings in computational effort.

In [52], [50] we have proposed an innovative new framework for optimization problems based on particle filtering (also called Sequential Monte Carlo method). This framework unifies and provides new insight into randomized optimization algorithms. The framework also sheds light on developing new optimization algorithms through the freedom in the framework and the various techniques for improving particle filtering.

The paper [23] considers a population-based Model-based Search (MBS), where a population of probabilistic models is maintained/updated and subsequently propagated from generation to generation. Unlike traditional MBS, one of the key questions in the proposed approach is how to efficiently distribute a given sample budget among different models in a population to maximize the algorithm performance. We formulate this problem as a MDP model and derive an optimal sampling scheme to adaptively allocate computational resources. In particular, the proposed sampling scheme assigns to each model a performance index to determine the quality of the model and samples the one that has the current best index. These performance indices are then further used in conjunction with a variant of the recently proposed cross-entropy (CE) method to update the current population to produce an improving population of models. We carry out numerical studies to illustrate the algorithm and compare its performance with existing procedures.

2.1.3 Applications

We have studied an application in data mining, specifically model-based cluster analysis involving maximum likelihood estimation. For large data sets with many clusters, the resulting log-likelihood function has many local optima, so traditional statistical techniques such as the popular Expectation-Maximization (EM) algorithm often fail to find the global maximum. We have recently applied our MRAS method clustering problems [20] and compared it with the EM method. Although the EM method is faster in converging to a local optimum, for problems with many local optima, it often fails to find the global optimum found by the other two methods, even with many restarts.

We have also made considerable advances in applying our sampling and model-based framework and related techniques to solving statistical problems. For example, we have proposed [29] two adaptive resampling algorithms for estimating bootstrap distributions. One algorithm applies the CE method and does not require calculation of the resampling probability weights via numerical optimization methods (e.g., Newton's method), whereas

the other algorithm can be viewed as a multi-stage extension of the classical two-step variance minimization approach. The two algorithms can be easily used as part of a general algorithm for Monte Carlo calculation of bootstrap confidence intervals and tests, and is especially useful in estimating rare event probabilities. We analyze theoretical properties of both algorithms in an idealized setting and carry out simulation studies to demonstrate their performance. In [30], we propose an adaptive importance resampling algorithm for estimating bootstrap quantiles of general statistics. The algorithm is especially useful in estimating extreme quantiles and can be easily used to construct bootstrap confidence intervals. Empirical results on real and simulated data sets show that the proposed algorithm is not only superior to the uniform resampling approach, but may also provide more than an order of magnitude of computational efficiency gains. We have introduced [43] a multi-step variance minimization algorithm for numerical estimation of Type I and Type II error probabilities in sequential tests. The algorithm can be applied to general test statistics and easily built into general design algorithms for sequential tests. Our simulation results indicate that the proposed algorithm is particularly useful for estimating tail probabilities, and may lead to significant computational efficiency gains over the crude Monte Carlo method.

Many clinical trials involve a sequential stopping rule to specify the conditions under which a study might be terminated earlier before its planned completion. The most important issue in the design stage is to determine the common operating characteristics such as Type I and Type II error rates, for which crude Monte Carlo simulation methods are widely adopted. However, it is well known that crude Monte Carlo may lead to large variabilities in resultant estimates and excessive waste of computational resources. In [31], we propose an efficient importance sampling approach for determining type I and type II error rates in both fully sequential and group sequential clinical trial designs with either immediate responses or survival endpoints. The approach is insensitive to the underlying statistics of interest, and can be easily built into a general algorithm to evaluate error rates, determine sample sizes, test statistical hypotheses, and construct confidence intervals.

Dose-response studies are routinely conducted in clinical trials to determine viable dose levels for newly developed therapeutic drugs. Due to safety, efficacy, and experimental design considerations, practical constraints are often imposed on (1) dose range (e.g. restricted dose range), (2) dose levels (e.g. the inclusion of placebo), (3) dose numbers (e.g. no more than four dose groups), (4) dose proportions (e.g. at least 20 percent of the subjects will be allocated to the placebo) and (5) potential missing trials. We propose [32] controlled optimal designs, that is, Bayesian multiple-objective optimal designs satisfying one or more of these practical constraints, for dose response studies. The resulting controlled optimal designs satisfying these realistic constraints can be readily adopted by the pharmaceutical researchers for optimal estimation of the parameters of interest such as the median effective dose level or the threshold dose level. We demonstrate our results and methodology through the logistic dose response model although our approach is viable for virtually any dose response model.

We have contributed to important advances the efficient allocation of a simulation budget. We have previously considered the problem of efficiently allocating simulation replications in order to maximize the probability of selecting the best design under the scenario in which system performances are sampled in the presence of correlation. For a general number of competing designs, an approximation for the asymptotically optimal allocation is

obtained, which coincides with the independent case derived previously in the limit as the correlation vanishes. An allocation algorithm based on the approximation is proposed and tested on several numerical examples. We have proposed [19] an optimal computing budget allocation (OCBA) method to improve the efficiency of simulation optimization using the CE method. In the stochastic simulation setting where replications are expensive but noise in the objective function estimate could mislead the search process, the allocation of simulation replications can make a significant difference in the performance of such global optimization search algorithms. The OCBA approach proposed here improves the updating of the sampling distribution by carrying out this allocation in an efficient manner. Numerical experiments indicate that the integration of OCBA with the CE method provides substantial computational improvement.

In [6], [21] we have presented a new sampling-based algorithm for solving stochastic discrete optimization problems. The algorithm solves the sample average approximation (SAA) of the original problem by iteratively updating and sampling from a probability distribution over the search space. We show that as the number of samples goes to infinity, the value returned by the algorithm converges to the optimal objective-function value and the probability distribution to a distribution that concentrates only on the set of best solutions of the original problem. We then extend the algorithm to solving finite-horizon MDPs, where the underlying MDP is approximated by a recursive SAA problem. We show that the estimate of the recursive sample-average-maximum computed by the extended algorithm at a given state approaches the optimal value of the state as the sample size per state per stage goes to infinity. The recursive algorithm for MDPs is then further extended to finite-horizon two-person zero-sum Markov games (MGs), providing a finite-iteration bound to the equilibrium value of the induced SAA game problem and asymptotic convergence to the equilibrium value of the original game. The time and space complexities of the extended algorithms for MDPs and MGs are independent of their state spaces.

In another application of simulation optimization, we have studied [16] the problem of determining the optimal control limits of control charts, which requires estimating the gradient of the expected cost function. Simulation is a very general methodology for estimating the expected costs, but for estimating the gradient, straightforward finite difference estimators can be inefficient. We demonstrate an alternative approach based on smoothed perturbation analysis (SPA), also known as conditional Monte Carlo. Numerical results and consequent design insights are obtained in determining the optimal control limits for EWMA and Bayes charts. The results indicate that the SPA gradient estimators can be significantly more efficient than finite difference estimators, and that a simulation approach using these estimators provides a viable alternative to other numerical solution techniques for the economic design problem.

2.2 Simulation-Based and Sampling Methods for Markov Decision Processes

We define an MDP $\{X_i, i = 0, 1, ..., T\}$ on state space $\mathcal{S}$ and action space $\mathcal{A}$ (cf. e.g., [1, 5]). In period (stage) i, the MDP in state $X_i \in \mathcal{S}$ takes action $a_i \in \mathcal{A}$, incurs cost $C_i(X_i, a_i, \omega_i)$, where ω_i denotes the stochastic element (e.g., random number), and then

transitions according to

$$X_{i+1} = f_{i+1}(X_i, a_i, \omega_i),$$

where $f_i(x, a, \cdot)$ denotes the (stochastic) transition function in period i for action a taken in state x. For notational simplicity, we have not made state and action spaces period dependent.

The objective is to find a feedback control **policy** $\pi \equiv \{\pi_i(x)\}_{i=0}^{T-1}$, which is a sequence of decision rules specifying the action a_i taken when in state x in period i, that minimizes an expected cost function, usually either finite horizon total cost, finite horizon discounted total cost, infinite horizon average cost, or infinite horizon discounted total cost. This proposal focuses on the discounted total cost setting, both finite and infinite horizon, and we define the **value function** associated with a policy and initial state:

$$V^{\pi}(x) = E\left[\sum_{i=0}^{T-1} \alpha^i C_i(X_i, \pi_i(X_i), \omega_i) \middle| X_0 = x\right], \tag{2}$$

where α is the (one-period) discount factor and T could be infinite, under the assumption that the limit is then well defined. As stated earlier, the chief context is the setting in which simulation is required to generate the system dynamics (state transitions) and/or period costs.

We begin by defining some familiar quantities:

$$\begin{aligned} Q_i(x, a) &= \text{(expected) cost-to-go (Q-function) in period } i \text{ for action } a \text{ taken in state } x \\ &\qquad \text{and optimal actions taken henceforth;} \\ V_i(x) &= \text{optimal value function in period } i \text{ for state } x. \end{aligned}$$

Then we have the usual Bellman optimality equation [1, 38]:

$$V_i(x) = \inf_a \left\{E\left[C_i(x, a, \omega_i) + \alpha V_{i+1}(f_{i+1}(x, a, \omega_i))\right]\right\}, \tag{3}$$

written here in two-part form:

$$Q_i(x, a) = E\left[C_i(s, a, \omega_i) + \alpha V_{i+1}(f_{i+1}(x, a, \omega_i))\right], \tag{4}$$

$$V_i(x) = \inf_a Q_i(x, a). \tag{5}$$

An optimal policy in period i will be denoted by

$$\pi_i^*(x) \in \arg\inf_a Q_i(x, a), \quad i = 0, ..., T-1, \quad x \in \mathcal{S}. \tag{6}$$

When the policy is stationary, the subscript/argument i will be dropped. In the infinite horizon stationary case, (3) takes the following form:

$$V(x) = \inf_a \left\{E\left[C(x, a, \omega) + \alpha V(f(x, a, \omega))\right]\right\}, \tag{7}$$

and we will assume there exists an optimal stationary policy such that

$$\pi^*(x) \in \arg\inf_a Q(x, a), \ x \in \mathcal{S}.$$

Input: stage $i < T$, state $x \in X$, $N_i > 0$, other parameters. (For $i = T$, $\hat{V}_T^{N_T}(x) = V_T^{N_T}(x) = 0$.)

Initialization: algorithm parameters; total number of simulations set to 0.

Loop until total number of simulations reaches N_i:

- Determine an action $\hat{a}$ to simulate next state via $f(x, \hat{a}, \omega)$, $\omega \sim U(0, 1)$.
- Update the following:
  number of times action a has been sampled $N_{\hat{a}}^i(x) \leftarrow N_{\hat{a}}^i(x) + 1$,
  Q-function estimate $\hat{Q}_i^{N_i}(x, \hat{a})$ based on $C(x, \hat{a}, \omega)$ and $\hat{V}_{i+1}^{N_{i+1}}(f(x, \hat{a}, \omega))$,
  the current optimal action estimate (for state x in stage i),
  and other algorithm-specific parameters.

Output: $\hat{V}_i^{N_i}(x)$ based on Q-function estimates $\{\hat{Q}_i^{N_i}(x, a)\}$.

Figure 1: Adaptive multi-stage (AMS) sampling framework

Traditional methods of policy iteration, value iteration, and variants based on linear programming all suffer from the curse of dimensionality. Furthermore, the transition function f_i is generally not known in closed form (note that in traditional MDP formulations, it is expressed in terms of explicit *transition probabilities* assumed given), but may be generated by a complicated stochastic simulation model, so in such a setting, the traditional methods are not directly applicable.

2.2.1 Adaptive Multi-stage Sampling

Adaptive multi-stage (AMS) sampling algorithms [3, 4] provide procedures for accurately and efficiently estimating the optimal value function under the constraint that there is a finite number of simulation replications to be allocated per state per period. These algorithms adaptively choose which action to sample as the sampling process proceeds, based on the estimates obtained up to that point, providing value function estimators that converge to the true value asymptotically in the number of simulation replications allocated per state. These algorithms are targeted at finite-horizon MDPs with large, possibly *uncountable*, state spaces but smaller *finite* action spaces.

Letting $\hat{V}_i^{N_i}(x)$ denote the estimate of the optimal value function $V_i(x)$ based on N_i simulations in period (stage) i, the objective is to estimate the optimal value $V_0(x_0)$ for a given starting state x_0. The approach will be to optimize over actions, based on the recursive optimality equations given by (4) and (5). The latter involves an optimization over the action space, so the main objective of the approaches is to adaptively determine which action to sample next. The chosen action will then be used to simulate the one-period costs $C_i(x, a, \omega_{i,j}^a)$ and next state $f_{i+1}(x, a, \omega_{i,j}^a)$, which are used to update the estimate of $Q_i(x, a)$ denoted by $\hat{Q}_i^{N_i}(x, a)$, which in turn determines the estimate $\hat{V}_i^{N_i}(x)$. Figure 1 provides a generic algorithm outline for the adaptive multi-stage sampling framework.

Specifically, $Q_i(x, a)$ is estimated for each state x and action $a \in \mathcal{A}(x)$, where $\mathcal{A}(x)$ is the set of admissible actions in state x, by a sample mean based on simulated next states and

rewards:

$$\hat{Q}_i^{N_i}(x,a) = \frac{1}{N_a^i(x)} \sum_{j=1}^{N_a^i(x)} \left[C_i(x,a,\omega_{i,j}^a) + \alpha \hat{V}_{i+1}^{N_{i+1}}(f_{i+1}(x,a,\omega_{i,j}^a)) \right], \tag{8}$$

where $N_a^i(x)$ is the number of times action a has been sampled from state x in period (stage) i ($N_i = \sum_{a \in \mathcal{A}(x)} N_a^i(x)$), and the sequence $\{\omega_{i,j}^a, j = 1, ..., N_a^i(x)\}$ contains the corresponding random numbers used to simulate the one-period costs $C_i(x,a,\omega_{i,j}^a)$ and next states $f_{i+1}(x,a,\omega_{i,j}^a)$. Note that the number of next-state samples depends on the state x, action a, and stage i.

In the general framework that estimates the Q-function via (8), the total number of sampled (next) states is $O(N^T)$ with $N = \max_{i=0,...,T-1} N_i$, which is independent of the state space size. One approach is to select "optimal" values of $N_a^i(x)$ for $i = 0, ..., T-1$, $a \in \mathcal{A}(x)$, and $x \in X$, such that the expected error between the values of $\hat{V}_0^{N_0}(x)$ and $V_0(x)$ is minimized, but this problem would be difficult to solve. We have developed two algorithms. Both construct a sampled tree in a recursive manner to estimate the optimal value at an initial state and incorporate an adaptive sampling mechanism for selecting which action to sample at each branch in the tree. The upper confidence bound (UCB) sampling algorithm chooses the next action based on the exploration-exploitation tradeoff captured by a multi-armed bandit model, whereas in the pursuit learning automata (PLA) sampling algorithm, the action is sampled from a probability distribution over the action space, where the distribution tries to concentrate mass on ("pursue") the estimate of the optimal action. The analysis of the UCB sampling algorithm is given in terms of the expected bias [3], whereas for the PLA sampling algorithm we provide a probability bound [4].

2.2.2 Population-Based Randomized Methods for MDPs

Population-based randomized algorithms for MDPs directly search the *policy space* to avoid carrying out an optimization over the entire action space at each policy iteration step, and update a *population* of policies as in genetic algorithms (GAs), using appropriate analogous operations for the MDP setting. The hope is that a population-based approach provides robustness, similar to GAs and scatter search in deterministic combinatorial optimization. The literature applying evolutionary algorithms such as GAs for solving MDPs is relatively sparse (e.g., [11, 34]). We have developed new population-based algorithms in [8, 27].

The first key feature of the algorithms in [8, 27] needed to ensure convergence to an optimal policy is an *elitist policy* that has a value function at least as good as the best value function in the previous population, i.e., $\widehat{\pi} \in \Pi_s$ is an elitist policy for $\Lambda \subset \Pi_s$ if $\forall\, \pi \in \Lambda$,

$$V^{\widehat{\pi}}(x) \leq V^{\pi}(x) \;\forall\, x \in \mathcal{S}.$$

If $\widehat{\pi}^k$ denotes an elitist policy for generation k, then it has the following monotonicity property:

$$V^{\widehat{\pi}^{k+1}}(x) \leq V^{\widehat{\pi}^k}(x) \;\forall\, x \in \mathcal{S}.$$

We use the term "elitist" (single policy) to distinguish it from "elite" (set of candidate solutions) in traditional GAs. The other key feature is an *action selection distribution* that generates mutations of policies to explore the policy space. An *action selection distribution* $\mathcal{P}_x$ for state $x \in \mathcal{S}$ is a probability distribution over the action space $\mathcal{A}$. The monotonicity

Input: population size $n > 1$, initial population Λ_0, action selection distribution $\mathcal{P}_x \; \forall \, x \in \mathcal{S}$, other policy mutation parameters.

Initialization: Set iteration count $k = 0$.

Loop until Stopping Rule is satisfied:

- Generate an **Elitist Policy** $\widehat{\pi}^k$ based on Λ_k.
- **Exploration**: Generate $(n-1)$ policies $\{\tilde{\pi}^1, \ldots, \tilde{\pi}^{n-1}\}$ via mutation operators and sampling from action selection distribution $\mathcal{P}_x$.
- **Next Population Generation:** Λ_{k+1}=$\{\widehat{\pi}^k, \tilde{\pi}^1, \ldots, \tilde{\pi}^{n-1}\}$, i=1, ..., $n-1$.
- $k \leftarrow k + 1$.

Output: $\widehat{\pi}^k$ an estimated optimal policy.

Figure 2: Population-based evolutionary framework for MDPs.

property of the elitist policy and the exploration property of the action selection distribution ensure that the algorithms converge to a population in which the elitist policy is an optimal policy. A description of a general framework for the population-based evolutionary approach is provided in Fig. 2, where $\Lambda_k \subset \Pi_s$ denotes the kth generation population of policies and $n = |\Lambda_k| > 1$ is the constant population size.

We have developed two algorithms: evolutionary policy iteration (EPI) and evolutionary random policy search (ERPS). These algorithms are especially targeted to problems where the state space is relatively small but the action space is extremely large, so that the policy improvement step in policy iteration (PI) becomes computationally impractical. They eliminate the operation of maximization over the entire action space in the policy improvement step by directly manipulating policies to generate an elitist policy. In earlier work under AFOSR support, we developed EPI [8], which uses a technique called *policy switching* [7] to generate an elitist policy from a set of given policies, with a computation time on the order of the size of the state space.

Evolutionary Random Policy Search

Evolutionary Random Policy Search (ERPS) is an enhancement of EPI that improves upon both the elitist policy determination and the mutation step by solving a sequence of sub-MDP problems defined on smaller policy spaces. As in EPI, each iteration of ERPS has two main steps: (i) generating an elitist policy, but using policy improvement with cost swapping (PICS) instead of policy switching, and (ii) exploring the policy using a "nearest neighbor" heuristic along with sampling of the entire action space.

Policy Improvement with Cost Swapping

ERPS splits a potentially large MDP problem into a sequence of smaller MDPs, to extract a convergent sequence of policies via solving these smaller problems. For a given policy population Λ, if we restrict the original MDP (e.g., costs, transition probabilities) to the subsets of actions $\Gamma(x) := \{\pi(x) : \pi \in \Lambda\} \; \forall \, x \in \mathcal{S}$, then a sub-MDP problem is induced from Λ as $\mathcal{G}_\Lambda := (\mathcal{S}, \Gamma, f, c)$, where $\Gamma := \bigcup_x \Gamma(x) \subset \mathcal{A}$. Note that in general $\Gamma(x)$ is a multi-set, which means that the set may contain repeated elements; however, we can always discard

the redundant members and view $\Gamma(x)$ as the set of admissible actions at state x.

Given a nonempty finite subset $\Lambda \subset \Pi_s$, a policy π_{pics} generated by *policy improvement with cost swapping* (PICS) with respect to the sub-MDP $\mathcal{G}_\Lambda$ is given by

$$\pi_{\text{pics}}(x) \in \arg\min_{a\in\Gamma(x)} \Big\{ E\big[C(x,a,\omega)\big] + \alpha E\left[\bar{V}^\Lambda(f(x,a,\omega))\right] \Big\}, \tag{9}$$

where $\bar{V}^\Lambda(x) = \min_{\pi\in\Lambda} V^\pi(x)\ \forall\, x \in \mathcal{S}$. PICS is a variation of the policy improvement step in policy iteration performed on the "swapped cost" $\bar{V}^\Lambda(x) = \min_{\pi\in\Lambda} V^\pi(x)$, hence the name "policy improvement with cost swapping". Note that the "swapped cost" $\bar{V}^\Lambda(x)$ may not be the value function corresponding to any single policy in Λ. However, the following result shows that the elitist policy generated by PICS improves any policy in Λ.

Theorem 2 *Consider a nonempty finite subset $\Lambda \subset \Pi_s$ and the policy π_{pics} generated by PICS with respect to $\mathcal{G}_\Lambda$ given by (9). Then, for all $x \in \mathcal{S}$, $V^{\pi_{\text{pics}}}(x) \le \bar{V}^\Lambda(x)$. Furthermore, if π_{pics} is not optimal for the sub-MDP $\mathcal{G}_\Lambda$, then $V^{\pi_{\text{pics}}}(x) < \bar{V}^\Lambda(x)$ for at least one $x \in \mathcal{S}$.*

According to this theorem, in each step of ERPS, the elitist policy $\widehat{\pi}^k$ generated by PICS with respect to the current sub-MDP $\mathcal{G}_{\Lambda_k}$, as given by (9), improves any policy in Λ_k. Thus, the new population Λ_{k+1} contains a policy that is superior to any policy in the previous population. Since $\widehat{\pi}^k$ is directly used to generate the $(k+1)st$ sub-MDP , the desired monotonicity property follows by induction.

Corollary 1 *Under ERPS, for all $k \ge 0$, $V^{\widehat{\pi}^{k+1}}(x) \le V^{\widehat{\pi}^k}(x)\ \forall\, x \in \mathcal{S}$.*

The computational complexity of each iteration of PICS is approximately the same as that of policy switching, because the policy evaluation step of PICS, which is also used by policy switching, requires solution of n systems of linear equations, and the number of operations required by using a direct method (e.g., Gaussian Elimination) is $O(n|\mathcal{S}|^3)$, and this dominates the computational complexity of the policy improvement step, which is at most $O(n|\mathcal{S}|^2)$.

Exploration via policy mutation and local nearest neighbor search

After PICS is used to generate an elitist policy, $n-1$ other policies $\tilde{\pi}^i,\ i = 1, ..., n-1$, are generated by the following procedure:

With probability q_0 (exploitation)
    choose action $\tilde{\pi}^i(x)$ in neighborhood of $\widehat{\pi}^k(x),\ x \in \mathcal{S}$, using "nearest neighbor" heuristic.
else (with probability $1 - q_0$) (exploration)
    choose action $\tilde{\pi}^i(x) \in \mathcal{A}$ according to $\mathcal{P}_x,\ x \in \mathcal{S}$.

These policies are then combined with the elitist policy to form the next generation.

For a metric space $\mathcal{A}$ with metric $d(\cdot,\cdot)$, we have the following convergence result.

Theorem 3 *Let π^* be an optimal policy with corresponding value function V^{π^*}, and let the sequence of elitist policies generated by ERPS together with their corresponding value functions be denoted by $\{\widehat{\pi}^k,\ k = 1, 2, \ldots\}$ and $\{V^{\widehat{\pi}^k},\ k = 1, 2, \ldots\}$, respectively. If (i) $q_0 < 1$, (ii)*

the action selection distribution satisfies for any $\ell > 0$, $\mathcal{P}_x(\{a|\ d(a, \pi^*(x)) \leq \ell,\ a \in \mathcal{A}\}) > 0,\ \forall\, x \in \mathcal{S}$, *and (iii) the transition matrix/function and one-stage cost function satisfy Lipschitz-type conditions, then*

$$V^{\widehat{\pi}^k}(x) \longrightarrow V^{\pi^*}(x) \quad \forall\, x \in \mathcal{S} \quad w.p.1.$$

2.2.3 Simulation-Based Approach to POMDPs

In a simulation-based approach to POMDPs, we have developed in [49, 51] a computationally viable and theoretically sound method for solving continuous-state POMDPs by effectively reducing the dimensionality of the belief space via density projections. The density projection technique is also incorporated into particle filtering (state estimation using simulations) to provide a filtering scheme for online decision-making. We have proved an error bound between the value function induced by the policy obtained by our method and the true value function of the POMDP, and also an error bound between the projection particle filter and the optimal filter. Finally, we have illustrated the effectiveness of our method through an inventory control problem.

3 Additional Research Progress

We have also made significant progress in the following areas:

- Perturbation analysis of a dynamic priority call center ([10]);
- Conditional Monte Carlo estimation of quantile sensitivites ([36]);
- Invited papers on simulation optimization and its applications [15], [35], [14], [9]..

4 Research Output

4.1 Journal Publications

- H.S. Chang, M.C. Fu, J. Hu, and S.I. Marcus. A Survey of Some Simulation-based Methods in Markov Decision Processes, *Communications in Information and Systems,* 7, 59-92, 2007.
- H.S. Chang, M.C. Fu, J. Hu, and S.I. Marcus. Recursive Learning Automata Approach to Markov Decision Processes, *IEEE Transactions on Automatic Control,* 52, 1349-1355, 2007.
- H.S. Chang, J. Hu, M.C. Fu, and S.I. Marcus, Adaptive Adversarial Multi-Armed Bandit Approach to Two-Person Zero-Sum Markov Games, *IEEE Transactions on Automatic Control,* forthcoming, 2010.
- M. Chen, J.Q. Hu, and M.C. Fu. Perturbation Analysis of a Dynamic Priority Call Center, *IEEE Transactions on Automatic Control*, forthcoming, 2010.

- C.H. Chen, D. He, M.C. Fu, and L.H. Lee. Efficient Simulation Budget Allocation for Selecting an Optimal Subset, *INFORMS Journal on Computing*, Vol.20, No.4, 579-595, 2008.

- M.C. Fu. What You Should Know About Simulation and Derivatives (Cover Story), *Naval Research Logistics*, Vol.55, No.8, 723-736, 2008.

- M.C. Fu, L.J. Hong and J.Q. Hu. Conditional Monte Carlo Estimation of Quantile Sensitivities, *Management Science*, Vol.55, No.12, 2019-2027, 2009.

- M.C. Fu, S. Lele, and T. Vossen. Conditional Monte Carlo Gradient Estimation in Economic Design of Control Limits, *Production & Operations Management*, Vol. 18, No. 1, 60-77, 2009.

- D. He, L.H. Lee, C.H. Chen, M.C. Fu, and S. Wasserkrug. Simulation Optimization Using the Cross-Entropy Method with Optimal Computing Budget Allocation, *ACM Transactions on Modeling and Computer Simulation*, in press.

- J.W. Heath, M.C. Fu, and W. Jank. New Global Optimization Algorithms for Model-Based Clustering, *Computational Statistics and Data Analysis*, in press.

- J. Hu, M.C. Fu, V. Ramezani, and S.I. Marcus. An Evolutionary Random Policy Search Algorithm for Solving Markov Decision Processes, *INFORMS Journal on Computing,* 19, 161-174, 2007.

- J. Hu, M.C. Fu, and S.I. Marcus. A Model Reference Adaptive Search Algorithm for Global Optimization, *Operations Research,* 55, 549-568, 2007.

- J. Hu and Z. Su. Efficient Error Determination in Sequential Clinical Trial Design, *Journal of Computational and Graphical Statistics,* 17, 925-948, 2008.

- J. Hu and Z. Su. Adaptive Resampling Algorithms for Estimating Bootstrap Distributions, *Journal of Statistical Planning and Inference,* 138, 1763-1777, 2008.

- J. Hu and Z. Su. Bootstrap Quantile Estimation via Importance Resampling, *Computational Statistics and Data Analysis,* 52, 5136-5142, 2008.

- J. Hu, M.C. Fu, and S. I. Marcus. A Model Reference Adaptive Search Method for Stochastic Global Optimization, *Communications in Information and Systems,* 8, 245-276, 2009.

- J. Hu, H.S. Chang, M.C. Fu, and S.I. Marcus. Dynamic Sample Budget Allocation in Model-Based Optimization, *Journal of Global Optimization,* forthcoming, 2010.

- J. Hu, W. Zhu, Y. Su, and W. K. Wong. Controlled Optimal Design Program for the Logit Dose Response Model, *Journal of Statistical Software,* forthcoming, 2010.

- H.G. Lee, A. Arapostathis, and S.I. Marcus, Necessary and Sufficient Conditions for State Equivalence to a Nonlinear Discrete-Time Observer Canonical Form, *IEEE Transactions on Automatic Control,* 53, 2701-2707, 2008.

- Z. Su, J. Hu, and W. Zhu. "Multi-Step Variance Minimization in Sequential Tests," *Statistics and Computing,* 18, 101-108, 2008.

- E. Zhou, M.C. Fu, and S.I. Marcus. Projection Particle Filtering for Dimension Reduction in Continuous-time POMDPs, *IEEE Transaction on Automatic Control,* forthcoming, 2010.

4.2 Refereed Proceedings or Book Chapters

- H. S. Chang, M. C. Fu, and S. I. Marcus. Adversarial Multi-Armed Bandit Approach to Two-Person Zero-Sum Markov Games, *Proceedings of the 46th IEEE Conference on Decision and Control,* December 2007.

- C.H. Chen, M.C. Fu, and L. Shi. Simulation and Optimization, *Tutorials in Operations Research*, Z.L. Chen and S. Raghavan, editors, INFORMS, 247–260, 2008.

- M.C. Fu. Variance-Gamma and Monte Carlo, *Advances in Mathematical Finance,* M.C. Fu, R.A. Jarrow, J.-Y. Yen, and R.J. Elliott, editors, Birkhauser, 21-35, 2007.

- M.C. Fu, C.H. Chen, and L. Shi. Some Topics in Simulation Optimization, *Proceedings of the 2008 Winter Simulation Conference,* Miami, FL, Dec. 7-10, 2008.

- J. Hu, M. C. Fu, and S. I. Marcus. A Model Reference Adaptive Search Method for Stochastic Optimization with Applications to Markov Decision Processes, *Proceedings of the 46th IEEE Conference on Decision and Control,* December 2007.

- J. Hu and H. S. Chang. A Population-Based Cross-Entropy Method with Dynamic Sample Allocation, *Proceedings of 47th IEEE Conference on Decision and Control,* 2426-2431, 2008.

- J. Hu and P. Hu. On the Performance of the Cross-Entropy Method, *Proceedings of the 2009 Winter Simulation Conference,* 459-468, 2009.

- U. Kuter and J. Hu. Computing and Using Lower and Upper Bounds for Action Elimination in MDP Planning, *Proceedings of the 7th Symposium on Abstraction, Reformulation and Approximation (SARA-07), Springer Lecture Notes in Computer Science (4612),* 243-257, 2007.

- A. Rawat, H. La, M. Shayman, and S.I. Marcus. Minimum Wavelength Assignment for Multicast Traffic in All-Optical WDM Tree Networks, *Proceedings of the 5th International Conference on Broadband Communications, Networks, and Systems (BROADNETS 2008),* London, UK, September 8-11, 2008.

- Y. Wang, M.C. Fu, and S.I. Marcus. A New Stochastic Gradient Estimator for American Option Pricing, *Proceedings of the European Control Conference,* Budapest, Hungary, August 23-26, 2009.

- Y. Wang, M.C. Fu, and S.I. Marcus. Sensitivity Analysis for Barrier Options, *Proceedings of the 2009 Winter Simulation Conference,* Austin, TX, Dec. 13-16, 2009.

- Y. Wang, M.C. Fu, and S.I. Marcus. Dynamic Pricing with Continuous Stochastic Demand, *Proceedings of the 48th IEEE Conference on Decision and Control,* Shanghai, China, Dec. 16-18, 2009.

- E. Zhou, M.C. Fu, and S.I. Marcus. Solving Continuous-State POMDPs via Projection Particle Filtering, *Proceedings of Eighth International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing,* Montreal, Canada, July 6-11, 2008.

- E. Zhou, M.C. Fu, and S.I. Marcus. A Particle Filtering Framework for Randomized Optimization Algorithms, *Proceedings of the 2008 Winter Simulation Conference,* Miami, FL, Dec. 7-10, 2008.

- E. Zhou, M.C. Fu, and S.I. Marcus. A Density Projection Approach to Dimension Reduction for Continuous-State POMDPs, *Proceedings of the 47th IEEE Conference on Decision and Control,* Cancun, Mexico, Dec. 9-11, 2008.

- E. Zhou, M.C. Fu, and S.I. Marcus. A Numerical Method for Financial Decision Problems Under Stochastic Volatility, *Proceedings of the 2009 Winter Simulation Conference,* Austin, TX, Dec. 13-16, 2009.

4.3 Authored Books or Monographs

- H.S. Chang, M.C. Fu, J. Hu, and S.I. Marcus. *Simulation-based Algorithms for Markov Decision Processes,* Springer-Verlag, 2007 (research monograph).

4.4 Edited Volumes

- M.C. Fu, R.A. Jarrow, J.-Y. Yen, and R. J. Elliott, editors, *Advances in Mathematical Finance,* Birkhauser, 2007.

4.5 Awards

- Michael Fu: Elected Fellow of the Institute of Electrical and Electronics Engineers (IEEE).

- Michael Fu: Elected Fellow of the Institute for Operations Research and the Management Sciences (INFORMS).

- Steve Marcus: Elected Fellow of the Society for Industrial and Applied Mathematics (SIAM).

- The paper "A Numerical Method for Financial Decision Problems under Stochastic Volatility," by Enlu Zhou, Kun Lin, Michael Fu, and Steve Marcus, won the Best Theoretical Paper Award at the 2009 Winter Simulation Conference (WSC), December 13-16, in Austin, Texas.

4.6 Ph.D. Students

- Pedram Fard, Ph.D., 2007, Univ. of Maryland, supervised by S. Marcus, M. Shayman, and H. La, "Dynamic Configuration of Network Topology in Optical Networks" (http://hdl.handle.net/1903/7412) (currently: Global Protocols)
- Huiju Zhang, Ph.D., 2007, Univ. of Maryland, supervised by M. Fu, "Three essays on stochastic optimization applied in financial engineering and inventory management," (http://hdl.handle.net/1903/6739) (currently: Fitch Ratings)
- Jeffrey Heath, Ph.D., 2007, Univ. of Maryland, supervised by M. Fu, "Global optimization of finite mixture models," (http://hdl.handle.net/1903/7179) (currently: Centre College)
- Scott Nestler, Ph.D., 2007, Univ. of Maryland, supervised by M. Fu, "Empirical analyses on federal thrift savings plan portfolio optimization," (http://hdl.handle.net/1903/7749) (currently: US Military Academy, West Point)
- Andrew Hall, Ph.D., 2009, Univ. of Maryland, supervised by M. Fu, "Simulating and Optimizing: Military Manpower Modeling and Mountain Range Options" (currently: US Military Academy, West Point)
- Matthew Reindorp, Ph.D., 2009, Univ. of Maryland, supervised by M. Fu, "Industrial Flexibility in Theory and Practice" (currently: Technical University of Eindhoven)
- Abraham Thomas, Ph.D., 2009, Univ. of Maryland, supervised by S. Marcus, "Learning Algorithms for Markov Decision Processes."
- Enlu Zhou, Ph.D., 2009, Univ. of Maryland, supervised by S. Marcus and M. Fu, "Particle Filtering for Stochastic Control and Global Optimization" (currently: Univ. of Illinois Urbana-Champaign)
- Ping Hu, Ph.D expected 2011, Stony Brook, supervised by J. Hu
- Yongqiang Wang, Ph.D expected 2011, Univ. of Maryland, supervised by M. Fu and S. Marcus
- Ranit Sengupta, Ph.D expected 2012, Univ. of Maryland, supervised by M. Fu and S. Marcus
- Kun Lin, Ph.D expected 2012, Univ. of Maryland, supervised by M. Fu and S. Marcus

References

[1] D.P. Bertsekas. *Dynamic Programming and Optimal Control, Vols. 1 & 2.* Athena Scientific, 1995.

[2] D.P. Bertsekas and J.N. Tsitsiklis. *Neuro-Dynamic Programming.* Athena Scientific, 1996.

[3] H.S. Chang, M.C. Fu, J. Hu, and S.I. Marcus. An adaptive sampling algorithm for solving Markov decision processes. *Operations Research*, 53:126–139, 2005.

[4] H.S. Chang, M.C. Fu, J. Hu, and S.I. Marcus. Recursive learning automata approach to Markov decision processes. *IEEE Transactions on Automatic Control*, 52(7):1249–1355, 2007.

[5] H.S. Chang, M.C. Fu, J. Hu, and S.I. Marcus. *Simulation-based Algorithms for Markov Decision Processes*. Springer, 2007.

[6] H.S. Chang, M.C. Fu, J. Hu, and S.I. Marcus. Adaptive adversarial multi-armed bandit approach to two-person zero-sum markov games. *IEEE Transactions on Automatic Control*, forthcoming 2010.

[7] H.S. Chang, R. Givan, and E.K.P. Chong. Parallel rollout for on-line solution of partially observable Markov decision processes. *Discrete Event Dynamic Systems: Theory and Application*, 15(3):309–341, 2004.

[8] H.S. Chang, H.-G. Lee, M.C. Fu, and S.I. Marcus. Evolutionary policy iteration for solving Markov decision processes. *IEEE Transactions on Automatic Control*, 50(11):1804–1808, 2005.

[9] C.H. Chen, M.C. Fu, and L. Shi. Simulation and optimization. *Tutorials in Operations Research*, 2008. in press.

[10] M. Chen, J.Q. Hu, and M.C. Fu. Perturbation analysis of a dynamic priority call center. *IEEE Transactions on Automatic Control*, forthcoming 2010.

[11] H. Chin and A. Jafari. Genetic algorithm methods for solving the best stationary policy of finite Markov decision processes. In *Proc. of the 30th Southeastern Symposium on System Theory*, pages 538–543, 1998.

[12] T.K. Das, A. Gosavi, S. Mahadevan, and N. Marchalleck. Solving semi-Markov decision problems using average reward reinforcement learning. *Management Science*, 45:560–574, 1999.

[13] P.T. de Boer, D.P. Kroese, S. Mannor, and R.Y. Rubinstein. A tutorial on the cross-entropy method. *Annals of Operation Research*, 134:19–67, 2005.

[14] M.C. Fu. Variance-gamma and Monte Carlo. In M.C. Fu, R.A. Jarrow, J.-Y. Yen, and R.J. Elliott, editors, *Advances in Mathematical Finance*, pages 21–35. Birkhäuser, 2007.

[15] M.C. Fu. What you should know about simulation and derivatives (cover story). *Naval Research Logistics*, 55(8):723–736, 2008.

[16] M.C. Fu, S. Lele, and T. Vossen. Conditional monte carlo gradient estimation in economic design of control limits. *Production & Operations Management*, 18(1):60–77, 2009.

[17] F.W. Glover. Tabu search: A tutorial. *Interfaces*, 20:74–94, 1990.

[18] D.E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning.* Addison Wesley, 1989.

[19] D. He, L.H. Lee, C.H. Chen, M.C. Fu, and S. Wasserkrug. Simulation optimization using the cross-entropy method with optimal computing budget allocation. *ACM Transactions on Modeling and Computer Simulation*, forthcoming 2010.

[20] J.W. Heath, M.C. Fu, and W. Jank. New global optimization algorithms for model-based clustering. *Computational Statistics and Data Analysis*, under review, 2009.

[21] M.C. Fu H.S. Chang and S.I. Marcus. Adversarial multi-armed bandit approach to two-person zero-sum markov games. In *Proceedings of the 46th IEEE Conference on Decision and Control*, pages 127–132, 2007.

[22] J. Hu and H.S. Chang. A population-based cross-entropy method with dynamic sample allocation. In *Proceedings of the 47th IEEE Conference on Decision and Control*, pages 2426–2431, 2008.

[23] J. Hu, H.S. Chang, M.C. Fu, and S.I. Marcus. Dynamic sample budget allocation in model-based optimization. *Journal of Global Optimization*, forthcoming 2010.

[24] J. Hu, M.C. Fu, and S. I. Marcus. A model reference adaptive search method for stochastic global optimization. *Communications in Information and Systems*, 8:245–276, 2008.

[25] J. Hu, M.C. Fu, and S.I. Marcus. A model reference adaptive search method for global optimization. *Operations Research*, 55(3):549–568, 2007.

[26] J. Hu, M.C. Fu, and S.I. Marcus. A model reference adaptive search method for stochastic optimization with applications to markov decision processes. In *Proceedings of the 46th IEEE Conference on Decision and Control*, pages 975–980, 2007.

[27] J. Hu, M.C. Fu, V. Ramezani, and S.I. Marcus. An evolutionary random search algorithm for solving Markov decision processes. *INFORMS Journal on Computing*, 19(2):161–174, 2007.

[28] J. Hu and P. Hu. On the performance of the cross-entropy method. In *Proceedings of the 2009 Winter Simulation Conference*, pages 459–468. IEEE, 2009.

[29] J. Hu and Z. Su. Adaptive resampling algorithms for estimating bootstrap distributions. *Journal of Statistical Planning and Inference*, 138:1763–1777, 2008.

[30] J. Hu and Z. Su. Efficient bootstrap quantile estimation. *Computational Statistics and Data Analysis*, 52:5136–5142, 2008.

[31] J. Hu and Z. Su. Efficient error determination in sequential clinical trial design. *Journal of Computational and Graphical Statistics*, 17:925–948, 2008.

[32] J. Hu, W. Zhu, Y. Su, and W.K. Wong. Controlled optimal design program for the logit dose response model. *Journal of Statistical Software*, forthcoming 2010.

[33] S. Kirkpatrick, C.D. Gelatt, and M.P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

[34] A.Z.-Z. Lin, J. Bean, and C. White III. A hybrid genetic/optimization algorithm for finite horizon Partially Observed Markov Decision Processes. Technical Report Technical Report 98-25, Dept. of Ind. and Oper. Eng., Univ. of Michigan, Ann Arbor, 1998.

[35] C.H. Chen M.C. Fu and L. Shi. Some topics in simulation optimization. In *Proceedings of the 2008 Winter Simulation Conference*, pages 27–38, 2008.

[36] L.J. Hong M.C. Fu and J.Q. Hu. Conditional monte carlo estimation of quantile sensitivities. *Management Science*, 55(12):2019–2027, 2009.

[37] W.B. Powell. *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. Wiley, New York, NY, 2007.

[38] M.L. Puterman. *Markov Decision Processes*. John Wiley & Sons, 1994.

[39] B. Van Roy and J.N. Tsitsiklis. Regression methods for pricing complex American-style options. *IEEE Transactions on Neural Networks*, 14:694–703, 2001.

[40] Y. Shen. *Annealing Adaptive Search with Hit-and-Run Sampling Methods for Global Optimization*. PhD thesis, Department of Industrial Engineering, University of Washington, Seattle, 2005.

[41] L. Shi and S. Ólafsson. Nested partitions method for global optimization. *Operations Research*, 48:390–407, 2000.

[42] R.L. Smith. Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research*, 32:1296–1308, 1984.

[43] Z. Su, J. Hu, and W. Zhu. Multi-step variance minimization in sequential tests. *Statistics and Computing*, 18:101–108, 2008.

[44] R.S. Sutton and A.G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, 1998.

[45] Z.B. Zabinsky. *Stochastic Adaptive Search for Global Optimization*. Kluwer, 2003.

[46] Z.B. Zabinsky, R.L. Smith, J.F. McDonald, H.E. Romeijn, and D.E. Kaufman. Improving hit-and-run for global optimization. *Journal of Global Optimization*, 3:171–192, 1993.

[47] Z.B. Zabinsky and G.R. Wood. Implementation of stochastic adaptive search with hit-and-run as a generator. In P. M Pardalos and H. E. Romeijn, editors, *Handbook of Global Optimization, Volume 2*, pages 251–273. Kluwer Academic, 2002.

[48] Q. Zhang and H. Mühlenbein. On the convergence of a class of estimation of distribution algorithm. *IEEE Trans. on Evolutionary Computation*, 8:127–136, 2004.

[49] E. Zhou, M.C. Fu, and S.I. Marcus. A density projection approach to dimension reduction for continuous-state pomdps. In *Proceedings of the 47th IEEE Conference on Decision and Control*, pages 5576–5581, 2008.

[50] E. Zhou, M.C. Fu, and S.I. Marcus. A particle filtering framework for randomized optimization algorithms. In *Proceedings of the 2008 Winter Simulation Conference*, pages 647–654, 2008.

[51] E. Zhou, M.C. Fu, and S.I. Marcus. Solving continuous-state pomdps via density projection. *IEEE Transaction on Automatic Control*, forthcoming 2010.

[52] E. Zhou, M.C. Fu, and S.I. Marcus. A particle filtering framework for randomized optimization algorithms: Edas, ce, mras, and more. *INFORMS Journal on Computing*, submitted for publication, 2010.

[53] M. Zlochin, M. Birattari, N. Meuleau, and M. Dorigo. Model-based search for combinatorial optimization: A critical survey. *Annals of Operations Research*, 131:373–395, 2004.